be protected, how you would then determine when you have a deterioration enough that we would be concerned?

MR. WINGES: Well, let me see if I can answer the last part of your question before I explain the chart. I think what I conclude from looking at these data, are that, once again, it's pretty much what I told you before, that the air quality in North Dakota in the Class I areas are about as good as it gets. That if your goal in doing the SIP recall is to try and get better concentration than this, you're probably barking up the wrong tree. You could probably shut down every power plant in the State of North Dakota and wouldn't get a lot better than this. It's my opinion. I haven't done that, but there's just not a lot of impact at these Class I areas.

What this shows, and somebody there on the staff correct me if I go off the deep end on what this says, but on this axis of the chart you have the 3-hour $SO2$ concentration. These are in ppb. All the numbers I've been presenting are micrograms per cubic meter, so pardon the unit change here.

MR. WITHAM: For clarification, the conversion is on the bottom of the chart.

MR. WINGES: You're right. Sorry. On this part of the chart you have frequency of occurrence. This is percent of the time. So these are in percent. So that number up there of 10 means 10 percent of the time. So, basically, you've got a concentration here of five parts per billion, which is about the detection limit of the instrument, and you're saying 10 percent of the time or so the concentration here at this monitor was at this level or at the detection level. As you move this way on the curve, you get to higher concentrations and, obviously, what you see is that those concentrations occur less often, less frequent.

Let's take this round curve here. This is a curve for the South Unit. The monitor, I'm assuming, Painted Canyon, and it was at Medora beforehand, through '85, and moved to Painted Canyon. They're saying that basically that here you get concentrations of 15 parts per billion or greater .01 percent of the time. That would be a fraction of .0001, and on this axis it looks like you have often in years that that would occur. So this dot here -- what does that say, .2 --

MR. SCHWINDT: .29.

MR. WINGES: -- .29. That says that --

looks like that that occurs, I think that would mean that would occur about once every four years, you would have a concentration of 15 parts per billion or greater. Am I reading that right? So essentially what this is saying is, that concentrations of any magnitude at the North Unit, Teddy Roosevelt North Unit -- I'm sorry -- South Unit are very infrequent. You just hardly ever have any concentrations above 15 parts per billion there. They're pretty good. Let's see, the North Unit is on here someplace, probably.

MR. WITHAM: It's on there twice.

MR. WINGES: Oh, yeah. So it appears -- yeah, what you've done here is, you've got the North Unit with the early part of the 1980s included when the concentrations were higher, and then you've got the time period from '84 through '98 where the concentrations were -- from the North Unit those early years where it was spiked up high are removed. That's this curve. It shows a similar pattern. This curve out here -- I guess it's this one, is for Teddy Roosevelt North Unit and, obviously, those high concentrations all occurred in that '80 through '84 period when the spikes that you saw in those curves earlier, we suspect that's oil and gas

sources that were close into the North Unit operating at that time.

I would just conclude from looking at that, that it confirms what I said earlier, and that is the concentrations in the park are pretty darn low, very low.

MR. WITHAM: I don't have any further questions.

MR. SCHWINDT: Any other questions?

MR. WINGES: Thank you very much.

MR. SCHWINDT: If we could, Bob, why don't we just take about a 10-minute break?

(Recess was taken at 2:31 p.m. to 2:43 p.m.)

MR. CONNERY: I would next like to introduce Bob Hammer, who many of you know. He has worked up here in North Dakota on many, many, many of your sources. I would dare to say that outside the Department there's probably nobody who knows the emission inventory and what's happened up here better than Bob Hammer. Bob, like many of the others here, is educated in meteorology. He's an expert in climate meteorology, monitoring, air quality monitoring, has been doing it for 20 years with Tetra Tech, ENSR, and other companies. And his education was at the South Dakota School of Mines and Technology in meteorology and Metropolitan State College.

I'm not going to go, again, through his project experience, but what Bob is going to talk about is what he does best and that has to do with the inputs to the modeling and going over what EPA -- what went into the EPA model and what went into the State model and what he thinks the best input to the model is going to be. Bob Hammer.

MR. HAMMER: Well, what I'd like to explain a little bit about today is we've been talking about the modeling process itself, and if we do get to the point that we have to do increment consumption analyses, we've been talking so much about modeling.

Kirk was able to talk a little bit about the modeling process in an overall context, and if you look at the modeling process and break it down, it has some key elements that are going to be the inputs to it, some of the main inputs that there are some items that will be gone over today, are emission inventory, which I will address. Mr. Paine, a little later, will do a little bit of some meteorology data and some effects of meteorology data, and, of course, the receptors, the

terrain and other information. All of that information ends up going into a computer to do some calculations and the output from that is that we end up getting some predicted impacts and that's what we're talking about, is those predicted impacts.

What is it we do with those predicted impacts once we've done that? Well, I'm going to show how a key element of that is how the emission inventory itself is handled, and I'm going to point out some things about the way the EPA did their emission inventory, as well as what the State did with theirs.

We've been talking a little bit about the PSD increment impacts and how really what the -- what we're talking about is a net change in air quality in a triggered planning area, and as it turns out, the whole State of North Dakota was triggered in 1977. And how is this compared to baseline conditions, and when you say "compared to," we've been talking a lot about applicable changes and what is the result of those applicable changes.

Well, we take the current emissions, run them through that model so that we can get an impact and we end up with a current impact. We then also need to take what the baseline conditions were over

the baseline emissions, run them through the computer model again and what we end up with is a baseline impact. And as we have pointed out several times, and as Kirk was able to go through some examples is, if you take that current minus baseline impact, you get your increment, and in that increment Kirk was able to show some negative numbers, as well as some positive numbers. You can show how the air quality has improved or how the air quality increment has been consumed. And when the number goes down, we call that expanding the increment itself.

What I'd like to do is go over some issues relative to EPA's graph modeling, as well as what the Department of Health did in their draft modeling report. If you look at what we're dealing with, and, really, some of the critical elements of this is, if you look down here at this number, this is what the level of the 3-hour Class I increment looks like for SO2 when we're talking about a NAAQS standard like this. So you can see how small changes in the increment are going to have potential significant effect in terms of how you handle that when you plug it into some calculations. And, of course, when you look at the 24-hour, you're dealing

with the same thing. You're dealing with such a small element, that those inputs to your calculations really need to be examined very closely and make sure they're done in the proper manner.

When you look at utility borders, and when I get down a little bit further, I'll talk a little bit about that, but a focus of what we've been looking at in large part these last few days is what have utility boilers done, how have they been handled. And if you look at how the State and EPA handled utility boilers, what they did was they based some emissions on CEM data, and so what were the current emissions and those were based on CEM data.

If you look at what the baseline emissions were and how they were handled by EPA, as well as the Department of Health, then that was based primarily on the AP-42 method or estimating emissions on the AP-42 calculations. Now, AP-42, I'm not sure how much we've gone into that, but AP-42 is a standard manual or document that EPA has that says, okay, if you don't have any real data, then we've gone out and over the years we've done a lot of testing and we've come up with mathematical equations that in general can be applied to industry and here's a way to calculate that. So if you don't have any real data, here's the way to do that.

And what we're doing is, we're taking these emissions that were based on CEMs for current data and applying them in the model. We're taking the emissions based on AP-42 and we're applying them in the model to get baseline impacts, and in the end what we're doing is, we're getting an increment impact that is going to be the difference between those. Now, again, a couple of different methods in that we're applying current emissions based on CEM or measured data, and in some cases EPA is working with different methods than the State is, but in both cases using measured data and AP-42 or calculated methods, because we simply didn't have the data back in the baseline period, so we have to come up with some method for estimating what those emissions were.

Now, if you look at what CEM data is like, really, CEM data because of the way that it is measured, it tends to be biased high. You can talk to any of the plant managers out here and the way that their CEMs operate and they will tell you that CEM data will come in and for the most part will be biased high. If you look at AP-42 calculations

relative to CEMs, what I'm going to show you is, that AP-42 calculations tend to be lower.

Now, we're talking about current emissions based on CEMs being biased high, baseline emissions on AP-42 being biased lower. Obviously, the result of that is that we're going to end up with an overestimate of increment consumption, and that is a concern and if you're talking about that small element of increment consumption that I had showed you in that graph, you've got to be careful with the way in which we're estimating emissions.

Just to give you an example of what it is that I was talking about, I've got two graphs on here. This is for data from Leland Olds Station Unit 1, and this is annual average emissions for the period of '96 through 2000, which at the time were the only years, five years that we had full years worth of data for. And the blue line shows what the annual emissions were from the CEM data, and the red or orange line here shows what the AP-42 calculations showed. And you can see how the AP-42 numbers underestimate what the CEM says, and if you look at it for Leland Olds Station, it's there and it's even more dramatic what the difference is between what AP-42 says -- again, this is all for

the same time period -- versus what the CEM says.

If you look at maximum hourly data, just take a peek at it, and, again, you deal with the same issue whether you're dealing with short-term data or you're dealing with annual data and you can see that we're, again, if you think back to that small element of increment consumption that we're dealing with and you can see that there can be a dramatic difference in the way that those emissions are estimated using these two methods.

What we looked at was how to deal with that and try to get into a case of doing a little more comparing apples to apples. Instead of apples to oranges, of just CEM data to AP-42, which our concern is you're overestimating increment consumption in that method, then for current data, if we go the way that EPA and Department of Health did and stick with CEM data, then we really need to come up with a baseline method that is an equivalent to CEM. We have shown in those graphs that AP-42, those standard calculations, they're not site-specific, do not do a great job of coming in and estimating what, in fact, that specific facility was like.

What we did was go out and do some detailed

statistical analyses to try to come up with an equation that would be site-specific to showing what the CEM equivalent was to try to basically duplicate the CEM data for that 1996 through 2000 time period, come up with an equation. Well, as it turned out, as many sophisticated ways as we tried to do this, it was a fairly simple equation. To some of you it may look familiar. Familiar very much to what AP-42, in fact, does. And despite trying some other more sophisticated ways of doing this, it did turn out that a simple factor of -- sorry about that -- 40.5 for Leland Olds Station Unit 2 and 36.0 times the sulfur percent and then the burn rate did the best to simulate what the CEM data was for that time period.

And just to demonstrate that, you can see the line here more closely simulates for Leland Olds Unit 1 what the annual data was and not underestimating in the fashion that AP-42 had done, and in the same way certainly more closely simulates what happened. Again, these are long-term emission rate data, and that's really how the analyses were done and the comparisons were. If you take that same data and, again, if you keep in mind that when EPA did their modeling, EPA did their modeling using short-term 90th percentile emission rates. And if you apply that data to what EPA did, you see that we're still largely underestimating what the CEM says for the most part, but at least we're performing better than what AP-42 did. And the same thing happens when you apply that data to short-term data. These are looking at maximums for the years. We're still underestimating what the CEMs are doing, but performing better than what AP-42 did. Now, this is just to demonstrate that there are some concerns relative to the emission calculation methods that were applied to what the Department of Health did and what EPA did in trying to estimate what those major source emission rates were.

Now, to just do an analysis or a summary of what EPA did in their emissions modeling inventory, when they looked at those major source 90th percentile emission rates, what EPA did was really a CEM to AP-42. They said that the current emissions were based on CEM and the baseline emissions were AP-42, when really the correct way to do that, more apples to apples, should be some method such as what I'm suggesting here, which is a CEM to CEM data. We've been able to show that there's better performance on a site-specific basis.

Another issue relative to what EPA did is, when they handled minor sources, they excluded minor sources. And, you know, being very familiar with the sort of sources that are out there, especially those around the park, having done a lot of work in North Dakota, also having looked at what Mr. Long presented yesterday relative to the potential influence of minor sources, EPA did recognize this in their report, but it wasn't concerned that they were out and any conclusions drawn from that analysis needs to be reconsidered, because those courses should be included.

Variant sources, as Mr. Connery pointed out relative to a legal issue and he dealt with this, the EPA had included variant sources in their emissions inventory for their modeling when, in fact, as Mr. Connery pointed out, sources like this for Class I impact analyses should be excluded. Normal operations analysis, when you figure out what you're going to use for your baseline data and you determine what your normal operations are, you start with those two baseline years, unless there's some reason at which you should change the normal operations year, and one of our presenters, Curt, will be going over some normal operations

information later on.

But what EPA did in their modeling analysis was stuck with 1976-77 and did AP-42 calculations for baseline, when, in fact, normal operations, if you look at the data for a number of facilities which the Department of Health did, and I believe did a very impressive document relative to coming up with baseline information, did, in fact, show the normal operations did vary from facility to facility.

Increment expanders, one thing that Kirk had pointed out was that the EPA did include increment expanders out there, but they did their modeling for all those other sources on this 90th percentile CEM and short-term emission rates for all the increment consumers, but then for some reason what they show is that annual emissions, for some reason, were used for increment expanders, so, you know, you can put it into thought, your short-term emission rates are higher, your annual emission rates are lower, you're going to be underestimating those and, of course, not getting a full effect from increment expansion for a short-term basis. When, in fact, you need to be consistent. If you are going to use annual emissions -- or short-term

emissions, you need to use short-term emissions for increment expanders as well.

And, you know, being very familiar with the situation here in North Dakota, I've done a lot of work in the Bismarck area, very familiar with the Mandan Refinery and some of the changes that have occurred at the Mandan Refinery over the years and how through the Title V process the emissions at the refinery have been reduced for a number of reasons and, in fact, there should be some increment expansion that should have been accounted for from the Mandan Refinery in EPA's modeling that they did not include, but, of course, the Department of Health being familiar with the situation themselves did include in their 2002 modeling, the Mandan Refinery.

If you look at what the Department of Health did and a few flaws that I think could be corrected to more closely simulate the emissions that are increment-affecting, be they expanding or consuming, is that when they handled the major utility sources, they also looked at more of a CEM to AP-42 comparison, when, in fact, as I pointed out earlier, is, you know, it's more accurate, I believe, to make a consistent comparison and a method such as the CEM to CEM that I presented would, in fact, tend to make things more realistic.

And normal operations, the Department of Health did go into a lot of detail relative to determining what normal operations would be for different facilities. However, I think there are some room for adjustments and reconsiderations in what the State has determined normal operations for those facilities should be, and I know that Curt will be getting into that a little bit as well.

So that summarizes my review of the emission inventory data, and I think there are some key elements in it that need to be considered to deal with such a fine-tuning -- such a small increment that we're dealing with right now. We have to make sure that the inputs going into the model -- if we have to model this thing, make sure that we're doing it the most appropriate way. And I think there are some things to reconsider in the modeling inventory.

MR. SCHWINDT: Thank you. You indicated that you believe the CEMs data is biased high. Do you know why?

MR. HAMMER: No, I don't know why it is biased high other than just some -- just some

information that I do have from talking with operators and managers and environmental managers at the facilities, that the way that the data has to be reported to EPA -- for one thing it tends to be biased high in that fashion. It is just built into it so that they purposely don't underestimate, is my understanding and I'm not real familiar with it, but if you do look at what the data itself shows in the graphs, we've been able to see that that does occur.

MR. SCHWINDT: Has there been any attempts to correct that bias on the numbers that are reported so that there is more accurate CEMs data that could be used?

MR. HAMMER: In the way that it's being reported? That would be best addressed by somebody that does that, so I'm not aware of that.

MR. SCHWINDT: Okay.

MR. BAHR: I didn't really follow how you came up with your new CEMs baseline, the figures that you used of the 40.5 times burn rate and such. How did you come up with those figures?

MR. HAMMER: How did we come up with those figures? What we did was, we took the CEMs data for those time periods and then looked at the number of factors that will affect that, including the sulfur

content, and really sulfur content, and I don't remember what other factors we looked at, but different things that could affect what the emission rates were, and we had some engineers and statisticians do some analyses to see what sort of relationships occurred between those different operating factors, including the fuel sulfur content and the emission rates that were being predicted by the CEMs, and after doing those statistical analyses, basically the best fit came out with that equation that I showed you.

MR. BAHR: It's really not feasible or possible for us to understand all that, right, without years and years of additional education, I'm assuming.

MR. HAMMER: Lots of mathematical analyses basically to see what kind of mathematical equation has a best fit to what the data actually does, and there are a lot of different variables, as the operators would tell you, that go into what effects what goes out the stack, but it's generally accepted sulfur content is one of those key factors, and we did, in fact, see that the sulfur content was over --

MR. BAHR: So, I guess, for us, basically,

all we can say is, based upon your graphs, it shows it's more close than just using the AP-42 process?

MR. HAMMER: Yes. That's right. And AP-42 is -- in many states, such as the State of Wyoming, if I were to go in and try to prepare a permit application and turn in an application and say that my emissions were estimated based on AP-42, the State of Wyoming would send me back out the door and say, no, AP-42 is not reliable, you have got to have better information than that. Although, in most places AP-42 is an accepted method and, you know, if you're going to go AP-42 to AP-42, that might be an acceptable way to do things, but we're comparing apples and oranges.

MR. SCHWINDT: Any other questions? Tom.

MR. BACHMAN: This is Tom Bachman, with the State Health Department. AP-42 indicates that SO2 emissions from lignite is highly dependent on the alkalinity of the coal. Did you take into account that alkalinity in doing your adjustment from AP-42 to CEM data?

MR. HAMMER: No, actually, Tom, what we didn't do is, we didn't go into it looking at we need to adjust AP-42. What we really did was go into it and do a pretty simple statistical analysis

391

to say what is a best-fit equation. So we didn't look at this as an adjustment to AP-42. And AP-42 does have a very similar equation. Instead, we looked at a lot of different equations. And I am very familiar with what you're talking about, about the alkalinity, and it says you can adjust what AP-42 does based on that alkalinity. Instead, we said, let's just find a site-specific factor and a best-fit mathematical approach.

MR. BACHMAN: Okay. Follow up to that, two of the facilities have obtained their coal from different mines than they were on the baseline date. Do you think that's a fair apples-to-apples comparison without taking into account the alkalinity in the coal?

MR. HAMMER: I wouldn't be able to say offhand, Tom, because from the statistical analyses that I saw most recently, I did not see a great variability and dependence on the alkalinity in the data that we had available for '96 through 2000.

MR. BACHMAN: Okay.

MR. SCHWINDT: Any other questions?

MR. WITHAM: Lyle Witham, Attorney General's office. One of the criticisms of the methodology that the State has used in its draft

392

modeling is that it may tend to underpredict highest concentrations. And I take it from what you're saying in your testimony is, that the apples-to-oranges comparison that you're talking about here using CEM, CEMS data to AP-42 will, in fact, underpredict -- or overpredict the concentrations in the model; is that a fair statement?

MR. HAMMER: That would be a fair statement, yes, because it will tend to -- relative to the current impact will be based on what the current emissions are, and if we're running higher on current emissions, the current impacts are going to run higher. Remember, the equation is the current impact minus the baseline impact. So if the baseline impact is running lower, then what you're subtracting is going to run lower, which, you know, just this equation, if you've got your current minus your baseline, this number is lower, your impact will be higher is the concern.

MR. WITHAM: In terms of interpreting the data and policy now, for the sake of argument, let's say that there may be some issues there. Are some of those balanced out by the overly conservative assumptions that you're suggesting that we're making by comparing CEMs to AP-42? Do you understand my

393

question?

MR. HAMMER: So are what balancing those out?

MR. WITHAM: The suggestion that the highest concentrations are underpredicted using average, are some of those compensated by the fact that the model is probably overpredicting because of some of the things you're talking about?

MR. HAMMER: I wouldn't consider them well-tempered. I wouldn't consider them balanced out well, because, you know, if you're going to -- it's all an issue of relativity. If you're talking about annual emissions maybe to try to smooth things out or temper things out, it's still a relative impact issue, what is the baseline relative to the annual.

MR. WITHAM: I have no further questions.

MR. SCHWINDT: Okay. Thank you. Any other questions?

MR. CONNERY: I would just like to mention -- to respond to some of the questions that were raised with Mr. Hammer. Mr. Melland is going to testify later, if you want to ask him a question about that. I'm not going to try. He has analyzed sodium content.

And I would just say that the reasons why

394

the State in its very careful analysis of what emissions were represented in the normal operation of all the sources source by source, obviously, a very careful analysis, rejected the same comment that Mr. Hammer just made based on the fact that we didn't do the analysis of sodium and we were using the factors that seemed to predict that more than 100 percent of the sulfur was converted in the burning of it, which is impossible.

Again, I would urge you to ask a question of Mr. Melland when he gets up here. The reason does have to do with continuous emission monitoring. I'm not sure that that Mr. Melland can answer, but we'll get someone else and we'll get it in writing to satisfy that concern. My understanding is, that the way the continuous emission monitoring works is when it breaks down, one of the ways that they encourage you to keep it operating 100 percent is you have to substitute values that are higher than it would have measured and that's how you get above 100 percent. These are very high temperature and very efficient conversions that we have.

Having said that, I would like to introduce Robert J. Paine of ENSR International, who is going to talk about Calpuff modeling, a subject that he knows better than anybody I've ever heard on this subject. And he's going to be talking about the meteorology that is involved in performing the kind of estimates that are required to do modeling well in this assessment that you're being asked to do.

Bob's experience, again, includes work for EPA. He happens to be an adviser to the American Meteorological Society and EPA and the working group that's working to revise the fundamental working horse model of EPA, ISC, and to develop a new model, AERMOD.

He has his degree in atmospheric science from the State University of New York at Albany. He has his master's in meteorology from the Massachusetts Institute of Technology. He has done -- he has 27 years of experience in doing this kind of evaluation. Several of his models have been adopted as guideline models. His work in developing models has in many cases been sponsored by EPA. He is deeply involved in exactly the kind of assessments that you're talking about here, which is long-range transport impact on Class I areas. So we have asked him to look at the modeling that has been done and to tell us what he thinks could be improved in that modeling. Bob.

MR. PAINE: I'm going to cover three areas. One is the history of Calpuff as it's evolved to be the recommended long-range transport model and its limitations. I'm going to look at the model evaluation study that was conducted by the North Dakota Department of Health to confirm and justify the manner in which the Calpuff model was run. And, finally, I'm going to suggest an alternative approach for running Calpuff for this modeling application.

First of all, some compliments to North Dakota. I think it has an excellent reputation in advanced long-range transport modeling. It's the first state I ever heard of that had run Mesopuff, and, in fact, it recognized and adopted Mesopuff more than 10 years ahead of the Interagency Work Group on Air Quality Modeling, which was a coordinated effort by the EPA and the Federal Land Managers, which would consist of the National Park Service, the Forest Service and the Fish and Wildlife Service. Before Calpuff was proposed, which was in the year 2000 by EPA to be a guideline model for long-range transport modeling, North Dakota was running that model, and I would like to commend the State for its conscientious effort in

this application.

The history of Mesopuff, was developed in the early 1980s, actually developed in my company, used to be ERT, worked for EPA. But EPA did not adopt it or any other model to be a long-range transport model. And long-range transport, as has been discussed before, is to transport distances beyond 50 kilometers, which has been used as a benchmark for when springline plumes are likely not to be realistic anymore and beyond that you had to have other techniques to more realistically define the trajectory of the plume. Seeing that there was no tool to address long-range transport modeling, these groups I mentioned before decided to coordinate their activities and try to come up with improved methods to do long-range transport modeling to assess the impacts of emissions at distant PSD Class I areas. Their first effort was a phase I approach to adopt Mesopuff as an off-the-shelf model in 1993, more than 10 years after North Dakota was using it.

And just to show you here, if the source is at -- if the source is at this asterisk here, Mesopuff emits a bunch of, if you're looking down on the earth here, circles called puffs, and they

enlarge with distance as they entrain ambient air and become more dilute, and their trajectory can be curved as the wind would bend their movement.

For those of you who do not recognize this part of the country where I come from, this is New England. This is a depiction of how you might show a wind field that would be able to more realistically than a straight-line model depict the movement of emissions. You can see that in each of these points there is an arrow and it shows the wind direction in speed by the length of the arrow. And this is a typical meteorological input to Mesopuff or Calpuff, and it's available at many levels and height. But this is a typical internal depiction of meteorology for those of you who are not familiar with how puff models work.

Back in the '80s we used decks of cards. This shows the -- I date myself, but this shows the type of technology we were using back when Mesopuff was originally done and it had many, many different sets of inputs that one had to take care of so it was very complicated. Calpuff is not really any easier. It's just that it's now on PCs and work stations.

But Mesopuff has limitations that the

Federal Land Managers and the EPA as part of the IWAQM group wanted to get beyond to a phase 2 model. It does not consider terrain effects. It only has two layers in the vertical so that the wind variation of height is of limited resolution. It has relatively primitive chemical transformation algorithms. So the IWAQM group planned to replace Mesopuff with Calpuff after it did some research into available models for a phase 2 approach to correct these shortcomings, and after further improvements and evaluations, EPA has proposed Calpuff finally, the first ever preferred long-range transport model that was ever proposed. And that was done on April 21st, 2000 in The Federal Register. And they're trying to basically go through all the comments and get the whole package together for maybe later this year to promulgate Calpuff.

EPA and IWAQM have conducted some limited evaluations of Calpuff for long-range transport databases. Now, these databases are expensive and rare so there's not too many of them. You have to go, obviously, far out and it takes a lot of manpower and equipment to do this. Many of the databases that were used in the evaluation have

observations out to about 100 kilometers from the release point. And, by and large, Calpuff has shown relatively unbiased predictions, which mean the predictions and observations aren't high or low relative to each other on average.

Here's an example of a layout of a Tracer study and there's a lot of symbols here, but the basic, most important things are the release points. And there's a row of receptors. In the distance here we see zero to 100 kilometers. So the farthest row of receptors would be maybe 100 kilometers, maybe 150 kilometers. And the reason why you have these all bunched together is because you want to know what the width of the plume is and you want to make sure you don't miss the peak. So you have a lot of these receptors lined up.

Now, obviously, you couldn't -- we wouldn't expect the State of North Dakota or any other state to put out this type of receptor network. This is a research-grade study and you can't maintain that very long before you run out of money. So it's usually very limited, maybe one or two hours of releases at most, which is why these data sets are limited in their duration and even their distance.

But what you can do with these measurements

is to define the plume concentration across from left to right at these monitoring arcs. And I'll show you an example in a minute. One of these experiments that IWAQM has researched and written about in their phase 2 document in 1998 was the Cross Appalachian Tracer Experiment, or CAPTEX. It had releases in Ohio and in Sudbury, Canada, and they had monitors all the way out to the Atlantic Ocean almost. And it was noticed in these -- in this particular experiment that Calpuff over-predicted by a factor of 3 to 4 at distances beyond 300 kilometers, and there was some concern about the ability of Calpuff to accommodate all the atmospheric phenomenon at such large distances. So IWAQM decided to place a cautionary note in their phase 2 document that indicates that the use of Calpuff beyond 300 kilometers has limitations due to wind shear, and I'll explain what that word means. It may not be evident to nonmeteorologists.

Wind shear over large puff dimensions. And I'll -- especially in the vertical, and I'll explain that with an illustration. Now, this modeling assessment involves transport distances approaching 200 kilometers from the major sources. If we take 100 kilometers as an overprediction ratio of 1

unbiased and 300 kilometers is 3 to 4, you might expect just by interpolation that an overprediction ratio could be expected on the order of 1 1/2 to 2 at the distances we're talking about in this study, and I'm going to show later on that that expectation is actually realized.

Here's a depiction of how this shear effect works. We're looking at a vertical cross-section of a puff, but I'm showing it as a series of particles. This comes out of a paper by Moran and Pilke, 1994. Let's say this is the shape of a puff or some sort of element of air with pollutants in it. At sunset and during the night, winds at different levels deform the puff, but this is not known to -- this is what really happens, let's say, and by the morning, early morning hours, it's really been bent in terms of its vertical straightness. It's very crooked. What happens during the morning is that as the sun comes up and the convective activity grows in depth, all this material is mixed to the ground uniformly, and when it is mixed to the ground, it covers a fairly large area because all this stuff gets down here and this material to the right is mixed way off to the right from this leftmost element.

But the Calpuff model, not realizing this wind shear has occurred, would mix it to the ground not just in this width as we see in panel H, but it would be only below the height or the width in panel E, very much a narrower and much more concentrated impact to the ground. And this is a limitation in Calpuff and you can imagine that the farther and the more time evolved in the transport of a puff, the more of these effects could be not taken into account.

In many evaluation studies IWAQM reported in its phase 2 document that Calpuff trajectories were in error, and this is by looking at graphs from the report, on the order of 20 percent and it could be less than -- 20 degrees, that is. More than 20 degrees. This follows Kirk Winges' talk that models, even very sophisticated models, cannot hit a bullseye, especially at the distances we're talking about, and due to the narrowness of the plume, a 20-degree miss can have significant error implications. Not only was the trajectory in error, but the plumes were generally observed to be too narrow relative to observations of the plume impact area.

I'm going to show you a diagram that illustrates this to try to show you. On the Y axis we have concentration. And on the X axis we have

sort of a horizontal degree along one of those arcs, so this is -- every little tick mark is 10 degrees, 10 degree arc. And we have in this sound line here labeled similarity dispersion is the predicted Calpuff concentration across an arc -- across an angular distribution. And the connected dots are the actual observations of concentrations. We note immediately that the Calpuff prediction is higher at the peak and narrower in its longitudinal extent than the observations, which is exactly what the problem is with its neglection or not being able to take into account the shear effects.

Here's another curve, by the way, using Pasquill-Gifford option, which is interesting because it was mentioned yesterday that the Pasquill-Gifford option might be assumed to lead to higher concentrations, but in this case it was lower so it's actually not always in one direction. But in this case the model, the Calpuff model, showed higher impacts, much higher observations, a displacement in angle and too narrow of a plume.

Now, IWAQM also looked into other more advanced meteorological data sets, because with observed data, such as airport data and balloon sets, all that Calpuff can do is do its best to

interpolate between the scattered observation points, but with meteorological prediction models there are some improvements available from this interpolation.

You can -- and I'm going to introduce a four-dimensional concept. You have three dimensions in space and one dimension in time, because in the inputs to the model you have to have winds -- on that plane we saw in New England you have a -- in one plane at one level you have many winds in X and Y, and you have several plains planes in vertical, that's three dimensions, and then you have to do it every hour, which is the fourth dimension. The input data will have many holes. The balloon ascents happen every 12 hours. The airports have -- and, of course, the balloon ascents are only sent up every so many hundreds of miles apart, and the airports are similarly located many tens of miles apart. Therefore, you have gaps in space and you have gaps in time. Therefore, you have many gaps in your four-dimensional arrangement to fill in for the model input, but a meteorological prediction model can optimize the filling of these gaps, and I'll show you why.

When you see the acronym FDDA, that means

four dimensional data assimilation. Try to take the available data and best fill in the space in time requirement for input to Calpuff. So the predictions of the winds between these observation points can be optimized with the laws of physics. For example, if you know that a front, cold front is passing an airport and you know its speed of progress, but one hour later you don't have an observation where the front was going to be, you know from the laws of physics, well, the wind must be shifted one hour's travel distance along. Whereas, if you didn't have that model, you wouldn't know that. So these models use laws of physics and progressions of wind changes to do better than just observations.

And there's also another acronym, mesoscale meteorological, which is MM, which is -- mesoscale is a sort of a distance scale of the order of 10 kilometers to hundreds of kilometers. And this is the type of distance scale we're trying to fill in missing observations, so these models tend to try to do their best to fill in missing winds in space and time. They're called mesoscale meteorological models. IWAQM has noted improved Calpuff performance with these FDDA-MM models and recognized their use, acknowledging that the data sets are costly. In view of this, since EPA allows, for example, less than five years and usually one year for on-site meteorological program at a single site, IWAQM recognizes that there is a corollary to using this enhanced meteorological data set. Therefore, in many cases in applications for PSD that I've done, the use of a single year of this enhanced meteorological data has been accepted in lieu of five years of just airport data because of its superiority, and it looks a lot like an on-site meteorological data program which EPA allows one year to be sufficient for an application. As I said, there is many applications where the -- for example, the 1990 MM4, which is version 4 of the meteorological model used by Penn State, that one year has been accepted in many PSD permit applications.

I was fortunate last year to be chairing over a committee meeting of the Airways Management Association Meteorology Committee, and we invited Joe Tikvart of the EPA's Office of Air Quality Planning and Standards to provide comments about the progress toward the promulgation of Calpuff, and here's some of his bulleted items. He noted that

EPA considers the Calpuff scientific basis to be sound, that Calpuff provides substantial transport and dispersion capabilities, as we've seen in just the wind field depictions. The accuracy is, quote, sufficient for long-range transport, which I take to be within a factor of 2 generally, to the limitations of the distance I mentioned, no more than 300 kilometers. Use of five years of data would be needed if you had only National Weather Services data. He mentioned that if you use a prognostic meteorological model, you could use one year and that's the same guidance as I mentioned for on-site meteorological data.

We've heard about the similarity option yesterday; that is, a nonIWAQM technique. However, EPA has determined that in this promulgation, it will recommend that this similarity option become the default, because improved performance has been noted and it's adopting the dispersion options of the other models. They're going to accept AERMOD, and they're going to throw out ISC, which uses PG, so it would be inconsistent for EPA to accept a similarity model like AERMOD for short-range and retain -- and not accept similarity for the long-range transport model. So similarity options

are likely to be accepted by EPA, and that is consistent with the way EPA and North Dakota have done their modeling so far, and I would recommend that they continue those options, because EPA is likely to recommend that option in its final promulgation.

Other improvements may be expected prior to promulgation, and I'm not sure what those are. They may be bookkeeping or other items. They have a beta test. EPA put out a beta test or a test version of Calpuff last year and any user input to that process may result in further changes.

I now have gone through the first element of my talk. I'm going now into how the model evaluation study by the North Dakota Department of Health, which was also used by EPA, looked to me and how I might reinterpret those results. I would like to say that North Dakota made a conscientious effort to take concurrent emissions for the year 2000 and meteorology and monitoring data and try to put together its best effort to look at how Calpuff is predicting in the way they ran the model. In an earlier evaluation study that I commented on on behalf of Minnkota, I noted that the hourly emissions from major sources were not used, so this

is an improvement.

The evaluation results indicated for two observation stations, Dunn Center and Theodore Roosevelt National Park South Unit, that there were reasonably unbiased modeling results. Just to show you the geometry that's involved here, these are -- the triangles indicate locations of major emission sources. This star is Dunn Center, the closer monitor, on the order of 100 kilometers from this group of sources. And the South Unit of Teddy Roosevelt National Park approaches 200 kilometers, maybe not quite, depending on which source you are talking about. Notice that there are no major sources that are in the direction to the south of this monitor here, which I'll refer to later on.

And I'm just recreating our -- the plots from the North Dakota report. Just to show you what this is, Kirk indicated or showed you other similar plots where you have observations on the X axis and predictions on the Y axis. And these predictions are unpaired in time and space. They would be much worse if they were paired. So you take the highest predicted and the highest observed, no matter when they occurred in the year 2000, and you plot that point up here, and you take the second highest

411

predicted and second highest observed and you plot that point, and so on.

And this is a factor of 2 overprediction, this line. This is a perfect prediction, and this is an underprediction by a factor of 2. And we see that most of the points are -- except for this one highest is a little bit underpredicted or overpredicted, and this would be about close to one and a half times the prediction over observation here. Going to the Dunn Center 24-hour, we see for the highest concentrations were unbiased or slightly overpredicting and then for just the lowest ones, we're underpredicting, but note that since the instrument threshold for detection is about five micrograms per cubic meter, it's two parts per billion, two tons, 2.62 is about 5.2 micrograms per cubic meter, that's about this line here, so we're dealing with -- by the way, I wouldn't be too concerned about these underpredictions because they're at the detection threshold of the instrument, anyway.

For Teddy Roosevelt South Unit 3-hour model predictions we are consistently overpredicting by about 1.5, treading upward above 1.5 to the highest value. Still within a factor of 2. And, finally,

412

for the 24-hour we are -- we crossed the line where the highest predictions are overpredicting. The lowest ones are underpredicting, but, again, we have this -- we have most of these underprediction values below the instrument threshold so not of sufficient concern. I did note the EPA was a little bit concerned about these underpredictions, but due to the fact that they are below the detection level of the instrument, I don't think that has to be worried about.

So these are what the North Dakota Department of Health provided and EPA recreated in their report, these last two figures. But I believe refinement is needed, because the total predicted concentration clearly stated in Section 9.2 of the EPA Modeling Guideline, 40 CFR Part 51, Appendix W, is that the total concentration is supposed to include the model portion plus the unmodeled portion, which is the regional background, and there is no regional background that I can tell in those predictions. That would mean that the predictions are understated.

Not only that, but oil and gas sources beyond 50 kilometers were not included because there are just a lot of them. That would tend to

413

underestimate the predictions. And the observations of zero were set to one part per billion, which is half of the two part per billion detection threshold. No correction was made to zero predictions so the observations were inflated in some cases, which would further tilt the prediction to observation ratio. So all these factors underestimated the prediction to observation ratios, but especially this first one where regional background was not included.

I made an effort to estimate what the regional background was likely to be, by -- as I noted before, there is no major sources to the south of the Teddy Roosevelt National Park South Unit. Looking at days during 2000, I got the hourly data from the monitor looking at days when the winds were clearly from the south. I reviewed the monitoring data for what the observations were then and they seemed to be running at at least one part per billion, combinations of zeros, ones and twos. For easterly winds I would expect, due to additional traffic, towns, and oil and gas sources that were not modeled, that you could have a regional background as high as one and a half parts per billion, four micrograms per cubic meter. Now, four

414

micrograms per cubic meter is still below the instrument threshold and lower than any regional background that any agency has allowed me to ever use for any modeling study. So I think that's a reasonably low estimate.

This is a picture of the Dunn Center monitor, by the way, and it's at a farm and may be subject to local sources that we have modeled. Teddy Roosevelt National Park we saw -- here's a truck on Interstate 94 here, and there's a main highway going right by the monitor. Those emissions aren't included in the modeling, so I would make the point that I think a four microgram per cubic meter on modeled background would be added as quite reasonable.

When you do that, you probably can't remember all these other curves, but we see that now Dunn Center 3-hour is ranging one and a half times predictions to observations. This point is still a little bit low at the very top. 24-hour is at least one and a half times as high as predictions to observations consistently, no crossover down here. Teddy Roosevelt South Unit, more than a factor of 1.5 overprediction on the 3-hour. The 24-hour is the most important because it's the most limiting.

415

The top concentrations are more than a factor of 2 too high. There is no underprediction at all.

It turns out that the farther monitor, Teddy Roosevelt South Unit, shows higher overpredictions than the Dunn Center one, consistent with the concept that the farther away you are from the sources, the more the effects of wind shear are neglected by the model, the more the plume is too concentrated and, therefore, you are going to get more overpredictions the farther away you go. It's very consistent with the expectations from the IWAQM study.

I would say this overprediction is cause for concern and I ask the question, would a different modeling procedure reduce this overprediction, which leads to the final part of my talk. I note that the data processing by the North Dakota Department of Health for 1990 to 1994 was optimized and done as carefully as possible, but what you just saw, the model overpredicts by a factor of 2 in some cases. The wind data coverage is quite limited. There is no upper air station between the sources and the monitor, the Class I areas and the closest Class I areas. There's only -- in terms of airports, there's Dickinson and

416

there's Bismarck, really. Many manual data replacements and adjustments had to be made to put this data set together.

I considered ways to enhance the way you could run the models, more horizontal resolutions, more vertical resolutions, trying more puff splitting, although in the end I found this wasn't very effective, and I think to this day that puff splitting is fairly ineffective in Calpuff.

This last item is, I think, most important, as I decided to try reaching the area of the four-dimensional data assimilation MM5 prognostic meteorological modeling data. For horizontal resolution, I'll go through those points one by one. We changed it from 10 kilometers to 3 kilometers to get better terrain effect resolution. Showing more realistic local winds that the terrain would induce, such as drainage flows. Also improving the resolution of the land use. John Vimont of the National Park Service says to me, well, I like 4 kilometers. Four kilometers is my favorite grade spacing and so we went one kilometer better than that.

This is what the terrain looks like with a 10-kilometer grid. These crosses are where the

417

receptors are. This is a modeling domain that goes close to -- let's see -- below 200 kilometers in this coordinate system to above 500, so that's almost 400 kilometers north/south, about 600 kilometers east/west. So that the sources will be over here. This is Teddy Roosevelt South, Elkhorn, Teddy Roosevelt North, Lostwood, and these are the Montana areas here. Okay. I'm going to go back and forth here. Keep your eye on that and I'm going to go one forward.

And 3 kilometers, we have much better resolution of these canyons and river values. Go back to one just to show you, then we go back to 10 kilometers, it's much fuzzier. As you would expect, the more resolution you have in your terrain, your grid, the better your terrain is going to look, the more realistic your terrain effects will be in the model.

Vertical layers, we -- that is, the EPA and the North Dakota Department of Health used eight layers to provide coverage in the vertical and I think that's not adequate enough over some of the puff vertical extent that I've seen, and I'll show you what that looks like. I think the use of 12 layers would give better coverage while providing

418

still reasonable demands on the computer resources. We have a tool called Caldesk. I'm going to dwell on this slide a little bit. Caldesk can tap into the model output and show you what it's doing.

This is a vertical. The Y axis is the vertical extent, and we're looking at a cross-section slice of the model and what's happening to a single source. It's sending out puffs. Here's where the source is, and we're looking -- this is left to right and this is in the vertical. These are older puffs that were emitted hours ago, and they were probably emitted during the day. When the sun went down, the atmosphere went to stable and these puffs are not growing very fast because the conditions are stable. We can see that these daytime puffs have a very large vertical extent, up to nearly 2,000 meters, but over most of their depth, the wind coverage is very inadequate. There is nothing over this depth here. We also noted an interesting phenomena where the winds are going from left to right at 1,500 or less meters here, 1,500 meters, and they're going the opposite way near the ground, which would lead to some puff deformations, which the model doesn't check for, anyway, but this is an example of wind shear.

Winds are going in different directions from the top of the puff to the bottom of the puff, bending the buff and eventually the puff would be mixed to the ground maybe more, but models won't be able to account for that. But I just want you to dwell on the fact that there is limited wind coverage over a large volume and extent of these puffs here. If you go to 12 layers, it's much improved. Generally speaking, the winds vary less with height as you go up, but still I think we have better coverage in this depiction of the 12 layers, so we decided to pursue this 12-layer model in the model we did after making this decision.

Now, the selection of the MM5 database; that is, meteorological model group 5 with four dimensional data assimilation. EPA in the 7th modeling conference, which is a public hearing for the proposed changes to the guideline, has recommended that forecast products from the National Weather Services, such as the Rapid Update Cycle, which many of you have probably never heard of, what I am going to refer to as the RUC, R-U-C. This is encouraged by the EPA and when you see some of the products from it, I think you'll understand why.

This model used to have data only every

three hours. In the year -- during the year 1999 hourly data became available in archive form from the National Weather Service. And it so happens that the year 2000, coincidentally, is the same year as the modeling archive -- as the model evaluation study. So those results are now subject to evaluation, the same evaluation North Dakota did. It turns out that the Rapid Update Cycle model is one of the most advanced weather-data gathering analysis systems available in the world, not available before so I can understand why North Dakota may not have stumbled upon this before now.

The weather data is updated every hour and the National Weather Service produces short-range forecasts after only 12 hours. It's very intensive, almost now-casting. It has a mixture of traditional sources of data and new sources of data. The traditional sources are surface stations every hour at airports and balloon soundings every 12 hours. Here's an example of a surface station at an airport. For example, it's only about 10 meters high, 33 feet, has an atmometer and a cup showing the wind speed and a wind vane. Obviously, even though it's -- these locations are widely distributed in space, there's more of these than

upper air stations, they only go up this high, so if your plume is much higher than this, it doesn't provide data any higher than 10 meters. This is a picture of a balloon release where the balloon has a long instrument train here with an instrument package below it. Released at widely separated points, Bismarck and Glasgow, Montana is all we've got. Once every 12 hours it goes up in the atmosphere and in one instant in time it gets the weather data, you wait 12 hours and you get it again. So we've got large gaps in time and space for this vertical data gathering for meteorology. This is what has been used for the 1990 to 1994 data.

The other nontraditional data sets available to the Rapid Update Cycle model can improve upon this greatly. There are three fairly new types of data; aircraft ascending and descending meteorology measurements. Airlines have been cooperating with the National Weather Service to have weather transponders on their planes. So if you're taking a trip on an airline and when the plane is ascending or descending, it's actually cooperating with the National Weather Service by sending that meteorological data as it's going up

and down to help those forecast models.

The most important item is actually the second one, the next generation radar, which I'll go into in a minute, gives you winds in a vertical space around Bismarck and other stations in the United States. And, finally, believe it or not, if there's a cloud moving out there, a satellite is looking at it and getting a wind from it.

This is a picture of a NEXRAD installation. There is one at the Bismarck airport. The dome is up here. It's up on stilts because it has to look out at a shelved angle and has to get above the local terrain. The one at Bismarck airport was installed in 1997. So, obviously, data from 1990 to 1994, did not include it. It's not available to the general public, anyway, but it's available to the National Weather Service. These are the locations in the United States for these NEXRAD radars. We see Bismarck here. I think Minot is up here. There's data coverage, and these circles represent the extent of coverage basically and where there's holes in the data, there's not observations, but, of course, this meteorological assimilation system will take diverse data and do the best it can with it. But, fortunately, since most of our emission sources are fairly close to Bismarck, data coverage is fairly adequate there.

Here's an example of an actual depiction of winds from the NEXRAD radar. You can get this on the Internet. And what is this futuristic looking thing? Well, let's say the winds are from the south/southeast. The green indicates that the winds are going toward the instrument here. The red indicates they're going away. And you have a depiction of the radio speed of the winds in terms of color coding available. And this is updated every 10 seconds, every hour, and it occupies a volume going all the way out here. So this is continuously available to the National Weather Service since 1997. And, of course, it also looks at clouds and rains, and so on, but it has a clear air mode where it can do winds. Never available before.

Cloud drifts, the cloud moving, the satellite will figure out from the drift velocity what the winds are, so if you see a lonely cloud in the sky, just feel happy that the National Weather Service is watching that cloud and getting the wind from it, improving its forecast for you. So these aircraft going up and down, the NEXRAD radar, and

the clouds drifting in the sky, believe it or not, those observations are going into this National Weather Service RUC model. They're automatically gathered and quality assured.

The previous one-hour forecast for the new hour are used if there's a data-poor region, so there might be data-poor regions. There's less of those than there are, obviously, without these, but it takes advantage of, again, the prognostic model moving fronts along into data-poor regions that can guess or estimate, initialize the winds for those data-poor regions.

The forecast and data analysis model accounts for land, water, mountain circulations, lake and sea breezes, snow cover, vegetation, very sophisticated forecast model. The National Weather Service has noted that it's seeing from this model relative to models it's had in the past, that there's better terrain definition, better analysis of small-scale texture variations, and it uses, by the way, the Mesoscale model version 5 for its analysis and forecast. It's embedded within the RUC model.

Wind analyses are better than previous models, as I mentioned. The frontal zones are

better depicted because it can predict observations of with -- observations of how the front went by to predictions of how they were going to go by. And it has better near surface wind results due to better surface characterization. It has a soil/vegetation model. It knows the snow and ice cover at all times. The version 2 model that is used, that I used in this year 2000 database, was fully implemented during the year 1999, so the year 2000 is really the first available full year. The data is analyzed at 40 levels in the vertical, extreme coverage of the vertical, and 40 kilometers in the horizontal. They are updated every hour, these analyses of the observed winds, but they are not saved by the National Weather Service. The National Weather Service is not in the business of air pollution forecasting or analysis. But there's a private company called Software Solutions and Environmental Services Company, SSESCO, that uses a no port satellite receiver to capture and archive these analyses hoping that someone will call and want to do a Calpuff run.

So what did we do? We called them. They gave us -- they worked out from their archives for the year 2000 a 40-kilometer grid of RUC data,

because now we're taking -- this RUC data is all that NEXRAD radar and the cloud drift winds and the aircraft goings and comings that never have been available before, and they put this into a 40-by-40 kilometer grid. But that's not all. They use a model to go to even finer resolution called the Advanced Regional Prediction System, or ARPS, and then there's another acronym, Data Assimilation System or ADAS, and you can get buried in these acronyms, but ARPS is used in specialized Mesoscale predictions, especially in Oklahoma for tornadoes. And I wish I could show you a simulation, but this is just a simulation of inflow to tornadic winds, and this is a very specialized model and if you could see the simulation, you'd see all sorts of rotations, and so on. This model is used to improve upon the RUC by SSESCO. So they take the archive RUC data to initialize the winds to a 40-kilometer resolution. They take surface observations and fine terrain down to 10 kilometers to derive wind data and 10 kilometer intervals for the MM5 data format, and they gave us this data set, 73 cells in east/west, 56 cells north/south, and this is, by the way, the surface stations used in that final step.

This is our grid that we used for running

427

Calpuffs, certainly every 10 kilometers. The 10 kilometers, that means at every one of these intersections there is every hour a vertical sounding of MET data. Just blows away anything that you could get from the 1990-1994 data, because in 1994 you have one point here in the vertical and one point up here once every 12 hours. Here you've got it every hour at all these points.

MR. SCHWINDT: Mr. Paine, I was just wondering, how long is your presentation going to go yet?

MR. PAINE: Sorry. Another half-hour at most. Should we take a break?

MR. SCHWINDT: Yeah, I think we probably should.

MR. PAINE: So I'll leave this up here for contemplation.

(A recess was taken from 4:00 p.m. to 4:10 p.m.)

MR. SCHWINDT: Okay. Thank you. Let's take a break for about 10 minutes.

MR. PAINE: Just to remind you, these points are every 10 kilometers, each of these intersections. So the Calmet analysis is 12 layers in the vertical, as I mentioned before, 3 kilometers

428

in the horizontal, which means we enhance the terrain in addition to the RUC and the ARPS terrain resolution. The surface and upper air data derived from the MM5 analysis was used in Calmet to make it run with these data sets.

And now I'm going to show you one two-day episode, and I shouldn't call it an episode. It's an event where there was an actual simulated plume transport toward the North Unit, but I want to show you the character of the winds. And non-meteorologists may not be able to appreciate how fine the detail is, but I'm going to show you. The way this works is, we have our modeling grade and we show the locations of the Class I areas; Teddy Roosevelt South Unit here, Elkhorn, just one little point here, North Unit here, Lostwood, Medicine Lake and Fort Peck. And the -- and the terrain is sort of shown by these colors here. The wind streamlines are shown down by these long lines with arrows, indicating that the winds are generally from the south at the start of this simulation, which the time up here is March 5th at midnight. The winds are going up to the northeast corner of this, towards the northeast, and I'm going to go one hour at a time showing an evolution of a system moving

429

through. Now I'm up to 3 o'clock. This visualization software is available with the data set. And what I'm going -- I'll sort of go through this and when I see -- here's an example of a circulation here that you couldn't possibly get with data with surface stations located far apart. This, by the way, the height of this wind field is near the surface.

I'm going to go now through -- here is 8 o'clock on the first of two days. We have here another circulation that would be impossible to see without surface stations located close enough so you could get the circulation in this fine detail, so this would be impossible to see with data sets of vintage 1990 to 1994 because the detailed winds in the scale we're seeing would not be resolvable by the surface stations. Here is another circulation. This is now moving down south, this counter-clockwise circulation here. It's a low pressure area. Now, we're at noon on the 5th of March, 2000. We see now that the situation is setting up in the afternoon of the 5th where the winds are going to now come from the southeast and start to evect plumes in the model toward these Class I areas. The circulation is developing to the south. It's

430

counterclockwise.

Now, I'm going through the rest of the day on the 5th of March, and we're seeing now that the winds are lining up very steadily from the east/southeast and the plumes, simulated plumes from the major sources would be evected towards the North Unit. Now I'm going to the end of the 5th of March, in the simulation and now going into the 6th of March, which is a day where the model predicted some elevated concentrations. And we see this on another fine scale circulation zone in the south part -- southwest part of the grid unresolvable with the types of separations of stations in the 1990 to 1994 years, but quite capable with the NEXRAD radar coverage in the year 2000.

Going now through the day of March 6th, every hour is a click. We're going to see this system here -- I'm going to go back one hour. Notice how this moves, this circulation moves abruptly north and keeps moving north, very wound-up circulation here, and it's going to -- we're going to see a plume simulation later on. But you can see that this now shifts, this circulation shifts now to the east of the Class I areas and the winds are now reversing direction suddenly and going from the northwest. This circulation is now moving rapidly to the southeast. This is the type of characterization of winds you can get from this RUC analysis. That would be much more detailed than you could ever get before. That's the end of that two-day simulation.

What I did was, I redid the evaluation with this new meteorological data, same observations, obviously, same emissions, just a different way to run the model, and I believe the results look better than with the previous run. Of course, adding the regional background I see at Dunn Center 3-hour. We had predictions before that were closer to 1.5 times the observed. Now we're very close, especially at the high predictions, to an unbiased model. The 24-hour, we are somewhat closer to the one-to-one line, but still slightly conservative so the model is still protective of air quality, which is encouraging, at the 100-kilometer distance range from major sources.

The Teddy Roosevelt 3-hour averages are -- they're closer to the one-to-one line, still overpredicting by maybe a factor of 1.3. And the -- you might remember that the 24-hour averages before were above a factor of 2 at the high end. Now,

they're on the order of 1.5. So we see an improvement or a decrease in the conservatism of the model. The overprediction ratio is reduced. The model is still overpredicting. You note that as you go further out from Dunn Center to Teddy Roosevelt South, the model overpredictions increase, as expected, from the distance dependency of the neglection of wind shear, but this data set is encouraging in that it's showing a better evaluation result, but still protective of air quality.

However, now here's a follow-up to what Kirk Winges was saying. With this best possible meteorological data can Calpuff now perform well on a paired in time and space basis, and the answer is no, it cannot, and I think that other researchers have found that it's an impossible goal, because turbulent eddies or little puffs of wind that cannot be measured or even modeled because it would just be too much of a task, cause slight plume trajectory deviations close to the source that forever after alter the trajectory of a plume, make good model performance paired in time and space an impossible goal, like it's impossible to make a weather forecast out to two weeks because a butterfly flapping its wings in Peking will eventually

contaminate the forecast.

It's just impossible, an impossible goal to have a model paired in time and space work out, especially at far, far distances. This is what you get, a gunshot, a scatter shot, a buckshot like Kirk was showing. You've got the highest -- this is prediction on the Y axis, observed on the X axis. Even though you have a number of predicted and observed points within the factor of 2 zone, this factor is an overprediction of 2. This middle line is a perfect model. This is an underprediction by a factor of 2 on the right. We have a considerable number of points that are high -- the predictions are high and these observations are zero, and vice versa, the observations are high and the predictions are nearly zero. So the goal of a model that can predict reliably in time and space I believe is unachievable and, therefore, the reliance upon such a technique to assess whether you can cancel out a baseline and a current prediction is unattainable and cannot be relied upon in any policy decision from a technical point of view, even with this data base.

I'd like to look at a comparison of the way North Dakota and EPA would run the year 2000 with

their eight-layer model, no MM5 data, versus the MM5 data to show why is the model predicting lower with the MM5 data for this event and other events. Same event as we just saw with the streamlines the second day. And this is a visualization tool called Caldesk, and Caldesk will show you the winds on the Calpuff grid at various heights. This height was at 375 meters. The length of the arrow at each grid point is proportional to the speed. Of course, it points in the direction of the wind. These stations, these triangles show you the surface stations, and these blue circles show you where the upper air stations are so that you can see that the coverage is quite sparse in the upper regions.

But I'm going to show you -- I'm going to go through a simulation where the time is now initialized to midnight on the 3rd -- I mean, on the 6th of March here, and we're going to see sort of translucent puffs emitting from this source. It can show you how puffs actually evolve inside the model. And I'm clicking one hour at a time. I'm at 3 o'clock in the morning now. The puffs are making a beeline toward the North Unit here from one of these modeled sources, and we can see that they are now starting to impact the North Unit at 8 o'clock in